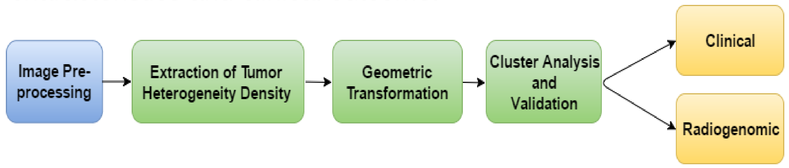
DEMARCATE: Density-based MRI Clustering for Assessing Tumor Heterogeneity in Cancer

A.Saha[1], S. Banerjee[2], S. Kurtek[1], S. Narang[3], J. Lee[3], G. Rao[4], J. Martinez[4], K. Bharath[5], A. Rao[3], V. Baladandayuthapani[6]

[1]Dept. of Statistics, The Ohio State University; [2]Operations Management and Quantitative Techniques Area, Indian Institute of Management Indore; [3]Dept. of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center (MDACC); [4]Dept. of Neurosurgery, MDACC; [5]School of Mathematical Sciences, The University of Nottingham; [6]Dept. of Biostatistics, MDACC.

INTRODUCTION

In this work, we examined MRI images from Glioblastoma Multiforme (GBM) patients to assess image-based tumor heterogeneity. Standard approaches are based on scalar summary measures of histograms (e.g. quantiles) that do not adequately capture the complete information in the voxel-level data. We introduce a novel technique, **DEMARCATE** (**DE**nsity-based **MA**gnetic **R**esonance image **C**lustering for **A**ssessing **T**umor h**E**terogeneity) to explore the entire tumor heterogeneity density profiles. Our analyses revealed two clusters of patients, with marked differences in tumor morphology, genomic characteristics and clinical outcome.



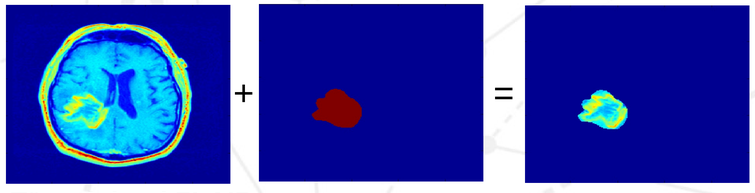
MOTIVATION and AIMS

- Glioblastoma Multiforme (GBM) is the most common and most aggressive malignant primary brain tumor.
- Tumor heterogeneity is a crucial area of cancer research wherein inter- and intra-tumor differences are investigated for assessing and monitoring disease progression.
- Imaging-based modalities allow for non-invasive techniques to assess tumor heterogeneity.

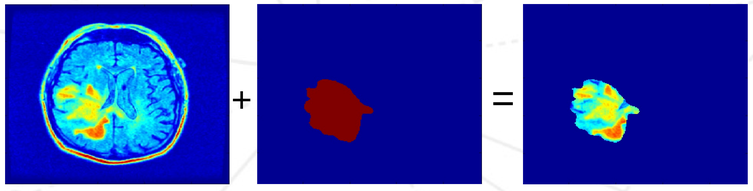
DESCRIPTION OF DATA

Number of patients considered for study: ***64***

T1-weighted post contrast



T2-weighted FLAIR

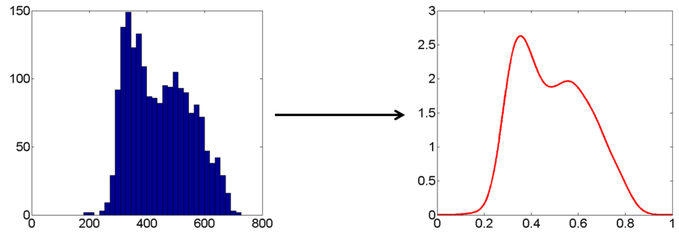


Covariates:
Tumor subtypes – ***Classical, Mesenchymal, Neural, Proneural***
Driver genes – ***DDIT3, EGFR, KIT, MDM4, PDGFRA, PIK3CA, PTEN***

METHODS

DEMARCATE STATISTICAL FRAMEWORK

Consider a smooth density representation of the histogram for pixel intensity values corresponding to the tumor.



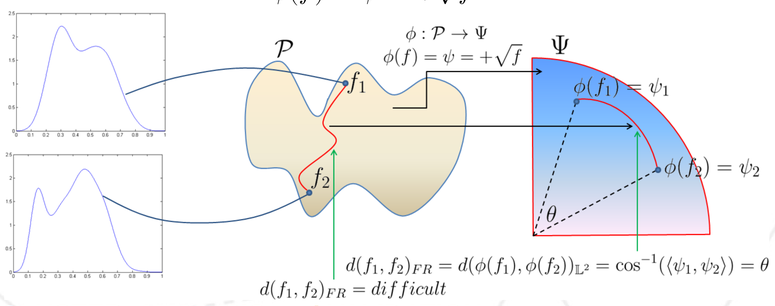
FISHER-RAO METRIC and REPRESENTATION SPACE OF PROBABILITY DENSITY FUNCTIONS

The *Fisher-Rao* (**FR**) metric on $\mathcal{P}$ is defined as:
for any point $f \in \mathcal{P}$ and the tangent vectors $\delta f_1, \delta f_2 \in T_f(\mathcal{P})$,

$$\langle\langle \delta f_1, \delta f_2 \rangle\rangle = \int_0^1 \delta f_1(t)\delta f_2(t)\frac{1}{f(t)}dt$$

Define *square-root transform* of probability density function:

$$\phi(f) = \psi = +\sqrt{f}$$



CLUSTER ANALYSIS and VALIDATION

We utilize an intrinsic version of the *k-means* clustering technique on Ψ. This approach partitions the space by minimizing the within-cluster sum of the squared distances (using the FR metric) to the assigned cluster centre.

For validation, consider a Bayesian version of *Fisher's Test*.

	Cluster 1	Cluster 2	TOTAL
A	y_1	n_1-y_1	$\mathbf{n_1}$
A'	y_2	n_2-y_2	$\mathbf{n_2}$

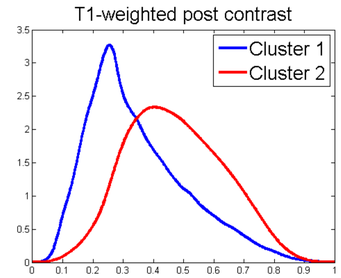
y_1 ~ Bin(n_1, θ_1)
y_2 ~ Bin(n_2, θ_2)

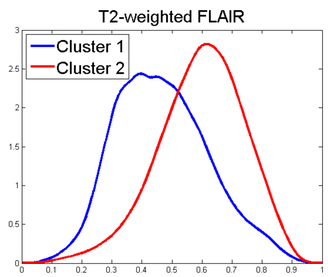
Let ***θ_1 ~ Unif(0,1)*** and ***θ_2 ~ Unif(0,1)*** be priors for proportions.
Posterior Distribution: ***$\theta_i \mid y_i$ ~ Beta (y_i + 1, n_i - y_i + 1)***, i = 1, 2.
Use Monte Carlo approximation to calculate *Enrichment Probability*:

$$\Pr[\theta_1 > \theta_2] \approx \frac{1}{M}\sum_{m=1}^{M} I(\theta_1^{(m)} > \theta_2^{(m)})$$

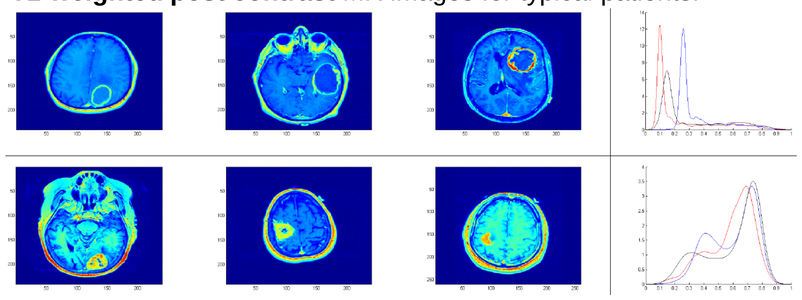
RESULTS

T1-weighted post contrast *Cluster 1*: *24*, *Cluster 2*: *40*
T2-weighted FLAIR *Cluster 1*: *30*, *Cluster 2*: *34*

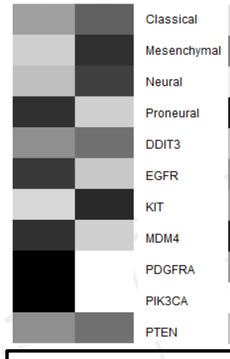




T1-weighted post contrast MR images for typical patients:



***T1* MRI**



***T2* MRI**

Cluster-wise survival times:

T1-weighted post contrast

	Cl 1	*Cl 2*	*Difference*
Mean	22.06	14.81	7.25
Median	17.00	12.95	4.05

T2-weighted FLAIR

	Cl 1	*Cl 2*	*Difference*
Mean	20.27	15.11	5.16
Median	16.60	13.45	3.15

CONCLUSION

1 = {high survival} + {Proneural} + {MDM4} + {young patients}
2 = {low survival} + {Classical, Neural} + {KIT, PTEN, DDIT3} + {older patients}

BIBLIIOGRAPHY & ACKNOWLEDGEMENTS

Just, N., 2014. *Improving tumour heterogeneity MRI assessment with histograms.* Br. J. Cancer 111 (12), 2205–2213.

VB and SB were partially supported by NIH grant R01 CA160736 and NSF DMS 1463233. AR and VB were also partially supported by the NIH through the University of Texas MD Anderson Cancer Center Support Grant (CCSG) (P30 CA016672).